The assessment of alcohol expectancies in school children: measurement or modification

Wiers, R.; Sergeant, J.A.; Gunning, B.

published in
Addiction
2000

DOI (link to publisher)
10.1046/j.1360-0443.2000.9557379.x

document version
Publisher's PDF, also known as Version of record

Addiction (2000) 95(5), 737–746

RESEARCH REPORT

The assessment of alcohol expectancies in school children: measurement or modification?

REINOUT W. WIERS,[1,2] JOSEPH A. SERGEANT[1] & W. BOUDEWIJN GUNNING[2]

[1]*Clinical Psychology, University of Amsterdam &* [2]*Child and Adolescent Psychiatry, Academic Medical Centre–University of Amsterdam and Amsterdam Institute for Addiction Research, Amsterdam, The Netherlands*

Abstract

Aims. *Earlier research has suggested that measuring children's positive alcohol-related expectancies could have the undesirable side effect of increasing them. This has been reported for an instrument that only measured positive expectancies and used a puppet-reference. The present study investigated whether this increase was still found using an unbiased instrument. Further, it was tested whether the assessment method with puppets influenced children's expectancies.* **Design.** *Children were assigned randomly to respond on an unbiased expectancy questionnaire in one of two assessment conditions: with reference to a puppet (P) or without reference to a puppet (Q). One month later, children were again administered one of the two assessment conditions, resulting in four assessment orders (PP, PQ, QP, QQ).* **Setting and participants.** *Three hundred and ninety-five second- to fifth-graders were administered one of the two methods in their schools and 260 children were measured a second time, 1 month later.* **Measurements.** *A questionnaire measuring children's positive and negative expectancies was developed that could be administered with or without a puppet-reference.* **Findings.** *A large direct response-effect was found: in the puppet condition, children scored higher on positive but not on negative expectancies. A smaller indirect measurement-effect was found at borderline significance: children who had used the puppet method 1 month earlier had significantly stronger positive expectancies than children who had used the questionnaire earlier.* **Conclusions.** *The present results confirm earlier indications that measuring children's positive expectancies may have the undesirable side effect of increasing them. This can be avoided by measuring children's expectancies with an unbiased questionnaire without a puppet-reference.*

Introduction

In the past two decades, alcohol-outcome expectancies have emerged as powerful predictors of both alcohol use and abuse in adolescents and adults (Goldman, Brown & Christiansen, 1987; Christiansen *et al.*, 1989). Expectancies are believed to be socially learned before personal experience with alcohol (Goldman *et al.*, 1987;

Correspondence to: Reinout W. Wiers, University of Maastricht, Faculty of Psychology, PO Box 616, 6200 MD Maastricht, The Netherlands. Tel: +31 43 388 1935 (/1905); Fax: +31 43 3884196; e-mail: R.Wiers@psychology.unimaas.nl

Submitted 24th November 1997; initial review completed 20th February 1998; final version accepted 28th October 1999.

ISSN 0965–2140 print/ISSN 1360-0443 online/00/050737–10 © Society for the Study of Addiction to Alcohol and Other Drugs

Carfax Publishing, Taylor & Francis Ltd

Fossey, 1994). Accordingly, schoolchildren endorse expectancies before they initiate regular alcohol consumption themselves (Miller, Smith & Goldman, 1990; Kraus, Smith & Ratner, 1994). This is important because expectancies prospectively predict onset of alcohol use (e.g. Christiansen *et al.*, 1989), and early onset of drinking is a risk factor for subsequent problem behaviours including alcohol abuse (e.g. Johnson *et al.*, 1995; Pedersen & Skondal, 1998).

A first intervention study aimed at modifying children's positive expectancies was reported by Kraus *et al.* (1994). Of central importance to the present study was their report that children's positive expectancies increased significantly after 1 month in the control condition, "*in the absence of any alcohol-related intervention other than the stimulation provided by the CARE's administration*" (Kraus *et al.*, 1990, p.5, italics added; "CARE" refers to a method of measuring Children's Alcohol Related Expectancies, Miller *et al.*, 1990). This finding raises the suspicion that expectancy *measurement* with the CARE could have an unintended side effect of *increasing* children's positive expectancies. The present study investigated this possibility. It was hypothesized that two aspects of the CARE assessment may have produced this increase: a response bias as a result of measuring only positive expectancies [1] and the assessment method involving a puppet-reference.

The CARE's focus on positive expectancies was linked to the same focus in the adult questionnaire from which the CARE was indirectly derived [1]: the Alcohol Expectancy Questionnaire (AEQ, Brown *et al.*, 1980). The AEQ measures only positive expectancies because of the theoretical assumption that people drink alcohol for reasons of expected reinforcement (Goldman *et al.*, 1987). However, there are good reasons to include negative expectancies, especially in children. First, a general methodological guideline for the construction of questionnaires prescribes the use of positive and negative (or contra-indicatory) items to minimize a positive response-bias (e.g. Cronbach, 1990). Secondly, it has been demonstrated that the inclusion of negative expectancies significantly increases the prediction of alcohol consumption in adolescents and adults (e.g. Fromme, Stroot & Kaplan, 1993; Leigh & Stacy, 1993; Wiers *et al.*, 1997). Negative expectancies are important with respect to an individual's decision to refrain from drinking (Jones & McMahon, 1994) and are likely to be relevant with respect to the initiation of drinking (Wiers, Gunning & Sergeant, 1998). Thirdly, although few studies have assessed children's expectancies, related research demonstrated that children primarily have *negative* attitudes concerning alcohol (e.g. Casswell *et al.*, 1985; Fossey, 1994).

A second aspect that complicated the interpretation of the increase in children's positive expectancies with the CARE method was the use of a puppet-reference. This procedure was introduced to provide a concrete reference and to focus children's attention (Miller *et al.*, 1990). While these may be valid *a priori* reasons, puppet-aided assessments have been criticized in other domains (i.e. children's testimonies of sexual abuse). The reason is that the use of a puppet appears to increase the child's suggestibility and motivation to respond in line with the presumed intent of the adult interviewer (Ceci & Bruck, 1993). Since this research has been carried out in a different domain and has indicated that the child's responses may depend on sometimes opposing cognitive and social factors, it is difficult to make exact predictions as to the effect of a puppet-aided assessment on children's expectancies. However, these findings suggest that the reported increase in children's positive expectancies may have been due partly to the use of a puppet-reference.

The present study focused on the effect of the assessment of children's expectancies on their subsequent development. As indicated, we had two hypotheses to explain the reported increase of children's positive expectancies. The first was that the increase was due to a response bias caused by measuring only positive expectancies. It was judged undesirable to test this hypothesis directly by comparing a balanced and an unbalanced questionnaire, given the reasons mentioned for not only measuring positive expectancies in children. Instead, we developed a balanced questionnaire measuring an equal number of positive and negative expectancies (in Dutch) that could be administered with or without a puppet-reference. As a result of this, the first hypothesis could be addressed only indirectly: by testing whether children's positive expectancies would still increase when this balanced instrument was used twice with a puppet. The second hypothesis was that children's positive expectancies had increased due partly to the use of a puppet-reference. It was not clear from the literature whether such a "puppet-effect" would occur in the present domain and whether it would influence positive and negative expectancies alike. This was investigated by comparing the two methods of assessment (with or without a puppet) in a 1-month follow-up design.

Method

Participants

In the first part of the study, 395 children participated from six suburban primary schools in different parts of the Netherlands. Some of the participating schools permitted only one administration for reasons of time and organization. It was decided to allow these schools to participate in order to obtain a large sample for initial factor analyses. The children were heterogeneous with respect to ability levels. All participating classes were mixed. The children were distributed approximately equally with respect to gender and grades. For the subsample tested twice ($n = 260$, main analyses) the distribution was as follows: 71 second graders (42 boys and 29 girls), 82 third-graders (40 boys and 42 girls), 48 fourth-graders (27 boys and 21 girls) and 59 fifth-graders (27 boys and 32 girls). Eighteen per cent of the children were of non-Dutch descent (mainly Turkish and Moroccan parents), but had all been raised in the Netherlands and mastered sufficient reading ability. The distribution of these children was equal across the conditions.

Materials

The children's version of the expectancy questionnaire was derived from the Dutch expectancy questionnaire for adolescents and adults (Wiers *et al.*, 1997). The questionnaire consisted of an equal number of positive and negative expectancies. Negative items included were positive statements with a contra-indicatory meaning of a positive expectancy. In general, this is a better way to develop contra-indicatory items than to introduce negations, especially for children who have difficulties in understanding negations. It should be noted that this procedure does not lead to a full coverage of the domain of negative expectancies, but should be optimal in preventing response bias. To increase comprehensibility, some of the original adult items were reformulated in simpler terms. As in the CARE method, the expectancies concerned the expected effects of alcohol on people in general, in order to allow for an answer in the absence of personal experience with alcohol. A five-point scale was used, with each of the five responses (I don't agree, I agree a very little bit, I agree a little bit, I agree, I agree very much) spelled out after each item. All items referred to expected effects "after a few alcoholic drinks". After piloting, the QED-C (Questionnaire for Expectancies in Dutch—Children) consisted of 30 expectancy items, 15 positive and 15 negative.

The QED-C started with a short alcohol recognition test, consisting of two questions: "is there alcohol in a bottle of cola?" and "Is there alcohol in a bottle of beer or wine?" This was followed by feedback of the correct answers and an opportunity for the children to raise questions. This brief introduction was chosen because previous research has shown that young children are able to recognize alcohol with great accuracy and appear to master the concept of alcohol (Noll, Zucker & Greenberg, 1990; Fossey, 1994). The expectancy section of the questionnaire was then introduced with an example item, followed by the 30 expectancy items. It was stressed that there were no correct or incorrect answers and that the investigators were interested in the children's individual opinions concerning what alcohol does to people.

Two different administration methods were used: a puppet-method (as in Miller *et al.*, 1990) and a questionnaire method. In the *puppet-method* (P) each expectancy-item was read aloud, while referencing to the expected effects of alcohol on a hand puppet. The general introduction of the CARE instructions were translated into Dutch (Miller *et al.*, 1990, p. 345). The items of the QED-C were adapted in such a way that the reference changed from the effect on people in general to the effect of a few alcoholic drinks on the puppet. As in the CARE, both a male and a female puppet were used. In the *questionnaire method* (Q) children answered the same questions without reference to puppets. Note that items in the two methods had the same content and differed only with respect to their referent: "Does Mr/Mrs Jansen [name puppet] enjoy a party after a few alcoholic drinks?" (puppet) vs. "Do people enjoy a party after a few alcoholic drinks?" (questionnaire).

Procedure

In consultation with the directors of the participating schools, a passive consent procedure was employed. The same method has been used repeatedly in this area (e.g. Fossey, 1994; Kraus *et al.*, 1994) and results in a much higher response rate in comparison with active consent procedures (cf. Query, Rosenberg & Tisak, 1998). All children were allowed to participate.

On the first measurement occasion, each class was randomly subjected to one of the two assessment methods (P or Q). On the second occasion, children of a class were randomly assigned to one of the two methods (e.g. if a class received the puppet method first, on the second occasion about half the children again received the puppet method and the others received the questionnaire). The subsample who were tested twice ($n = 260$) may be categorized by the order of the two methods they received at time 1 (T1) and at time 2 (T2): 62 children received the puppet-method twice (PP-condition), 74 children received the questionnaire twice (QQ-condition), 47 children received the puppet-method on T1 and the questionnaire on T2 (PQ-condition) and 77 children received the reverse order (QP-condition). There were no significant differences between these subgroups with respect to gender-distribution and age.

Statistical analyses

As a prerequisite for a meaningful comparison across methods and subgroups (differentiated for age group and gender) a common factor model was tested using multi-group restrictive factor analyses (LISREL 8, Jöreskog & Sörbom, 1993). The goal of these analyses of the *measurement model* was to obtain a common factor model that could be applied across methods, age groups and gender. These analyses were performed with the entire sample measured at T1 ($n = 395$). Since the present study used a new expectancy instrument, the analyses commenced with an exploratory factor analysis. This procedure is not strictly confirmatory, but nor is it trivial, because it explicitly tests whether a comparison across methods and subgroups is possible (Jöreskog, 1993, see Wiers *et al.*, 1997). The analyses of the measurement model were followed by a series of multivariate analyses of variance investigating differences in expectancies between subgroups ($2 \times 2 \times 2$ MANOVA T1-data, $n = 395$, between-subject factors gender, age-group, method). Given the number of independent variables, it was decided to differentiate between two age-groups only (grades 2 and 3 vs. grades 4 and 5). The effects of the measurement-method used at T1 on expectancies at T2 were analysed with a repeated-measures MANOVA and with a between-subjects MANOVA of the T2 data, with the method used on T1 and on T2 as independent variables ($n = 260$). Finally, a multi-level analysis was performed to test the repeated-measures effect simultaneously at class and individual level (with MIXREG, Hedeker & Gibbons, 1996).

Results

Alcohol identification

Ninety-five per cent of the children answered that a bottle of cola does not contain alcohol but that a bottle of wine or beer does. Since the correct answer was provided for the children afterwards, all children were retained.

Expectancies: measurement model

Four items had given rise to questions concerning their meaning and were removed from the analyses. Three further items concerning cognitive–motor enhancement showed very skewed responses (no endorsement at all, as in the CARE, Miller *et al.*, 1990) and were also removed from the analyses.

The remaining 23 items were analysed for all subjects with complete data who participated at T1 ($n = 395$). For reasons of space, the analyses of the measurement model are reported briefly here. In a series of restrictive factor analyses models containing two, three and four correlated latent expectancy factors were compared. A four-factor model best fitted the data obtained with the questionnaire method. However, the same model could not be fitted to the data of subgroups that had received the puppet method. Since the main purpose of the paper was to compare expectancies *across methods*, further possibilities were investigated to obtain a measurement model that would fit across methods. As a solution, a model with a second-order factor structure was specified, with four first-order factors (sexual enhancement, excitement, antisocial behaviour and negative mood) and two second-order factors (positive and negative expectancies). Comparable second-order models have been proposed for adult and adolescent expectancies (Leigh & Stacy, 1993; Wiers *et al.*, 1997). This model fitted modestly: χ^2 (979) = 1448.6 $p < 0.001$, RMSEA = 0.060, CFI = 0.74 [2].

Table 1. *Expectancy items in the two measurement models*

After a few alcoholic drinks people [questionnaire]/Mr or Mrs Jansen [puppet] ...	
2nd Order factor: positive expectancies	2nd Order factor: negative expectancies
1st order factor: "sex"	1st order factor: "antisocial"
• talk about sex more easily	• become annoying while playing a game
• feel like making love	• become cheeky
• kiss more readily	• want to fight
* feel little like making love (-)	• say stupid things
1st Order factor: "excitement-fun"	1st Order factor: "negative mood"
• feel tough	• feel unattractive
• find it exciting to drink alcohol with friends	• see the future gloomy
• enjoy drinking together while watching TV	• feel insecure
• feel good after a hard day's work	* make a party become annoying
* forget their problems	* are bad at playing snooker
* enjoy a birthday-party	
* become quiet (-)	
* lose a bad mood	
* become funny while playing games	
* are not scared of anything	

Items beginning with * were not included in the reduced 14-item measurement model; items beginning with • were included in both measurement models. Hence, the reduced measurement model consisted of seven positive and seven negative items and the 23-item measurement model consisted of 14 positive (4 "sex", 10 "excitement-fun") and nine negative items (four "antisocial", five "negative mood"). Items ending with (-) were scored inversely.

Nine items showed poor reliability (squared multiple correlations below 0.10). These items could be removed without affecting the factor structure. The reduced model (with 14 items) showed a good fit across subgroups split with respect to age-group and method: χ^2 (338) = 418, $p = 0.0018$, cRMSEA = 0.050, CFI = 0.89, and split with respect to gender and method: χ^2 (338) = 439, $p < 0.001$, cRMSEA = 0.056, CFI = 0.87. The items of both models may be found in Table 1. Further details of the measurement models can be found in Wiers (1998) and are available upon request. The main point of the analyses of the measurement model is that is has been possible to fit a common factor model across methods and subgroups allowing for meaningful comparisons. The choice between the 23-item model and the reduced 14-item model is difficult: the reduced model fits the data better, but has the disadvantage that only 14 of the original 30 items are used, and that the item selection has been based on the T1 sample. We decided to present the results primarily for the 23-item model, and briefly for the reduced 14 item-model in case of different outcomes.

Differences in expectancies in relation to age-group, gender and method (T1, n *= 395)*

All two- and three-way interactions were non-significant ($p > 0.10$). Significant multivariate main effects were found for age-group, F (2, 386) = 5.0, $p = 0.007$, method, F (2, 386) = 19.3, $p < 0.001$ and a trend for gender F (2, 386) = 2.4, $p = 0.09$. In the absence of interactions the interpretation of the main effects is straightforward.

Method. Children who received the puppet method scored significantly higher than children who received the questionnaire method on positive expectancies, F (1, 387) = 36.3, $p < 0.001$, but not on negative expectancies ($p > 0.10$).

Age group. Older children scored significantly higher than younger children on positive expectancies, F (1, 387) = 10.1, $p = 0.002$, but not on negative expectancies ($p > 0.10$).

Gender. Boys scored significantly higher on negative expectancies than girls, F (1, 387) = 4.8, $p = 0.029$, but not on positive expectancies ($p > 0.10$). Outcomes with the reduced 14-item model differed only with respect to gender: boys scored higher than girls on both positive and negative expectancies ($0.01 < p < 0.05$). Again, much larger main effects were found for method ($p < 0.001$) and age group ($p < 0.01$).

Positive and negative expectancies after one month (T2, n *= 260)*

Test–retest reliabilities for each of the four or-

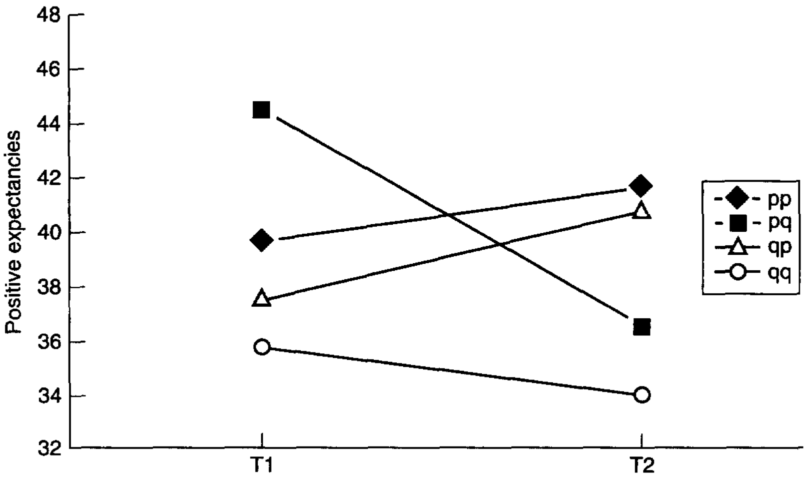

Figure 1. *Scores on positive alcohol expectancies at T1 and T2 as a function of the assessment methods used at the two measurement occasions (puppet—puppet condition; puppet—questionnaire condition; questionnaire—puppet condition; questionnaire—questionnaire condition).*

ders of methods were modest (but all significant, $p<0.01$) for positive expectancies: 0.54 (PP), 0.55 (PQ), 0.42 (QP) and 0.63 (QQ), and for negative expectancies: 0.46 (PP), 0.64 (PQ), 0.59 (QP) and 0.63 (QQ). When corrected for test-length, the PP reliability is similar to that of the CARE [3], and the QQ reliability is better.

With repeated measures MANOVAs, scores of the 260 children who had received two measurements were analysed in a mixed design, with three between-factors: gender, age-group and order of assessments (PP, PQ, QP, QQ), and one within-factor (time). We focus here on the time effects (between effects were reported above for the larger sample). All two- and three-way interactions of between-factors with time were not significant ($p>0.10$). The only between-factor that significantly interacted with time was order of assessment, F (6, 486) = 9.0, $p<0.001$. The order by time interaction was highly significant for positive expectancies, F (3, 244) = 17.9, $p<0.001$, but not for negative expectancies ($p>0.10$). Figure 1 shows the scores on the positive expectancies at T1 and T2 as a function of assessment methods at both measurement moments.

A large direct response-effect can be observed: children who received the puppet-method at T1 score higher than children who received the questionnaire at T1, and the same pattern is found at T2 (in line with the method effect described above). The next question is whether there is an additional effect of method used at T1 on expectancies measured one month later. There is an indication of such an effect, because positive expectancies appear to increase in the PP condition but not in the QQ condition. To test this hypothesis, the change in positive expectancies in the PP and QQ condition were compared in a repeated measures MANOVA (only PP and QQ conditions, $n=136$). The order by time interaction was significant, F (1, 134) = 7.8, $p=0.006$, with positive expectancies increasing significantly in the PP condition, F (1, 134) = 4.4, $p=0.039$, and decreasing at borderline significance in the QQ condition, F (1, 134) = 3.4, $p=0066$. Outcomes with the reduced 14-item model were in the same direction, but stronger: a significant order by time interaction, F (1, 134) = 12.3, $p=0.001$, with positive expectancies increasing significantly in the PP condition ($p<0.001$) but not in the QQ condition ($p>0.50$). The hypothesis of an indirect measurement effect was further tested by comparing the expectancies at T2 as a function of the assessment method at T1 and at T2 (2×2 between-subject MANOVA, $n=260$). No significant interaction was found between the methods used at both moments ($p>0.10$). The direct response-effect at time of measurement (here T2) was again highly significant, F (1, 256) = 32.3, $p<0.001$, with higher scores in the

Table 2. *Means and standard deviations of positive expectancies on T1 and T2 in the two measurement models in the original sample* (n = 260) *and after exclusion of four classes* (n = 205)

		Full model (23 items)				Reduced model (14 items)			
		T1		T2		T1		T2	
	Methods used	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Original sample ($n = 260$)	PP	39.8	8.2	41.8	8.6	7.7	2.7	9.4	2.7
	PQ	44.7	10.0	36.6	8.7	9.1	2.8	8.1	2.8
	QP	37.5	8.3	41.0	8.1	7.6	2.6	9.0	2.9
	QQ	35.8	8.0	34.1	8.7	7.7	2.4	7.7	2.9
Reduced sample ($n = 205$)	PP	41.3	8.8	42.5	8.4	8.2	2.9	9.8	2.6
	PQ	45.3	12.9	37.3	11.2	9.2	3.3	7.9	3.2
	QP	36.5	7.9	41.0	8.3	7.3	2.5	8.8	2.9
	QQ	35.6	8.4	34.0	9.0	7.6	2.4	7.6	2.9

PP = puppet-method at T1 and at T2; PQ = puppet-method at T1 and questionnaire method at T2; QP = questionnaire method at T1 and puppet-method at T2; QQ = questionnaire method at T1 and at T2.

puppet condition. The indirect effect of main interest here (positive expectancies at T2 as a function of assessment method at T1) was smaller, but significant, F (1, 256) = 2.8, $p = 0.047$ (one-tailed: hypothesized increase). In the reduced model, the same test was at borderline significance ($p = 0.07$). In Fig. 1, this effect can be observed by comparing the two pairs of conditions that used the same method on T2: the conditions that used the puppet-method at T1 score higher on T2, irrespective of method used at T2 (PQ > QQ and PP > QP). Hence, it appears that in addition to the strong direct response effect with children scoring higher on positive expectancies when a puppet is used, there is a smaller indirect effect of method used a month earlier which is at borderline significance.

From the graph another unexpected difference appeared: the scores of children who received the same instrument at T1 differed as a function of the instrument they were to receive later, especially the two subgroups that received the puppet method at T1 (PP and PQ). This was tested formally and confirmed (MANOVA expectancies at T1 as a function of method used at T1 and T2 which showed a significant interaction, $p = 0.002$). Whether this difference was the result of outliers at the level of the participants was tested first, but this was not the case. Another possibility was that something had gone wrong in the random assignment at T2 in some of the classes (remember that a class received one method at T1 and children were randomly assigned to one of the two methods at T2). This hypothesis was tested by comparing for each class the scores on positive expectancies at T1 as a function of method received at T2. In four of the 14 classes a significant difference was found. After removal of these classes there was no longer a significant difference at T1 between children who received the same instrument at T1 as a function of measurement method at T2 ($p > 0.10$). Table 2 shows the means of the positive expectancies in each of the four conditions, for the sample before and after removal of these four classes.

The analyses of the indirect measurement effect were repeated with the remaining 10 classes ($n = 205$). In the 23-item model, the change scores in the PP and QQ conditions (repeated measures MANOVA, $n = 110$) were different at borderline significance F (1, 108) = 3.4, $p = 0.07$, and they remained significant in the 14-item model, F (1, 108) = 7.3, $p = 0.008$. The main effect of method used at T1 on positive expectancies at T2 remained significant for the 23-item version, F (1, 201) = 3.12, $p = 0.04$ and close to significance in the 14-item version, F (1, 201) = 2.48, $p = 0.06$ (both one-tailed).

Given the indications of a randomization error in some of the classes, a final control analysis was performed with the remaining 10 classes, in which both the class-level and the individual level were taken into account (multi-level analysis with MIXREG, Hedeker & Gibbons, 1996). With this analysis the development of the positive expectancies in 1 month in the PP and the QQ condition were compared. Again a significant main effect of condition was found (puppet method higher, $Z = 3.6$, $p < 0.001$). The effect of main interest (the interaction of condition with time) was analysed by the multi-level effect of condition on the change scores. It was again at borderline significance for the 23-item measurement model, $Z = 1.8$, $p = 0.066$ and significant for the 14-item model, $Z = 2.7$, $p = 0.006$, with positive expectancies increasing in the PP condition. The outcomes of the multi-level analyses confirmed the analyses described earlier at the level of the individual: a large direct response effect and a smaller indirect measurement effect, significant or at borderline significance depending on the measurement model used. The similarity of the outcomes was not surprising given the intraclass correlations: $r = 0.066$, $p = 0.16$ for the main effect and $r = 0.0$, $p > 0.50$ for the interaction.

Discussion

This study investigated the suggestion from earlier research that children's positive alcohol expectancies could increase as a result of their measurement. Two aspects of the assessment procedure were hypothesized to produce such an undesirable effect: a response bias due to only including positive expectancies and the use of an assessment method with a puppet-reference.

The first assessment-factor (response bias) was investigated only indirectly here. No direct comparison was made with an instrument only measuring positive expectancies, given the good reasons not to use such an unbalanced instrument. However, the hypothesis could be tested indirectly: by testing whether positive expectancies would still significantly increase when a balanced instrument was used with a puppet-reference (PP condition). This was the case.

The second assessment-factor was investigated directly here by comparing two assessment methods (puppet vs. questionnaire) in a 1-month follow-up design. When discussing the differences between these assessment-methods, two effects have to be distinguished: a direct response-effect and a long-term measurement effect (effect of assessment-method used at T1 on expectancies 1 month later). The direct response-effect was large: participants scored much higher on positive expectancies when they were given the puppet method compared with the questionnaire ($p < 0.001$) in the absence of significant differences in their scores on negative expectancies. This large effect was independent of gender and age of the participants and occurred at both measurements moments. The second effect, an increase in positive expectancies as a result of the measurement method used at T1, was much smaller (at borderline significance). One may question what the "clinical significance" is of such a small effect. For two reasons we believe the effect is important, even though it is small. First, the direction of the effect is opposite to the effect any alcohol researcher ultimately has; an increase in positive expectancies as a result of the investigator's measurement is undesirable because positive expectancies predict an early onset of drinking with its negative consequences (see Introduction). The second reason is that this undesirable effect can so easily be avoided: all one has to do is to use a balanced questionnaire without a puppet-reference. The use of a puppet-reference appears to be both undesirable and unnecessary: the puppet method not only increases children's positive expectancies, but the psychometric properties of the puppet method were also inferior to those of the questionnaire (in terms of factor structure and test–retest reliability). The difference as a result of *both* factors (response bias and puppet-reference) can be inferred from the comparison of the repeated assessment of the CARE (positive expectancies increased with 9% in 1 month, Kraus *et al.*, 1994), with the repeated assessment of the questionnaire here: no increase at all.

One could speculate *how* the higher scores on positive expectancies in the puppet-method come about. It seems likely that the more group-orientated context of the puppet method plays a role. Even though investigators may try to minimize the interaction between the children, in the puppet-method children have to look up to see the puppet, which provides them with ample opportunity to see other children's reactions. This could also stimulate discussions afterwards, which might be a factor in the suggested increase in positive expectancies as a result of a previous assessment. It has been shown that social and motivational factors are of importance when a puppet is used (Ceci & Bruck, 1993). Further studies could be designed to find out how exactly this effect of the assessment method comes about. Alternatively, one may conclude from the present study that children's expectancies should be measured with an unbiased instrument without a puppet-reference.

Although the present study was designed primarily to investigate expectancy assessment in children, some more general conclusions concerning the development of children's expectancies may be drawn. Consistent with the earlier findings of Miller *et al.* (1990) and Kraus *et al.* (1994), the older age group (grades 4 and 5) scored higher on positive expectancies than the younger age group (grades 2 and 3). This study shows that this effect is found in the absence of significant differences between these age groups in negative expectancies. However, Query *et al.* (1998) found a different pattern: no age-effect on positive expectancies and a significant decrease in negative expectancies with age. A longitudinal study of the development of children's positive and negative expectancies is needed. This would also be interesting with respect to the emergence of expected cognitive and motor enhancement, which is not found in children (Miller *et al.*, 1990; this study), but predicts alcohol consumption in adolescents (Christiansen *et al.*, 1989; Wiers *et al.*, 1997). Perhaps this expectancy only develops after personal experience with alcohol (see Wiers *et al.*, 1998).

Some limitations of the present study should be mentioned. First, the instrument developed here does not cover the whole domain of alcohol expectancies. In particular negative expectancies are not fully covered because only contra-indicatory items of positive expectancies were included as negative items. Secondly, the test–retest reliabilities of the scales used were modest, although comparable with the original CARE. The effect of a modest reliability on comparisons of group means is that it lowers the power to detect a statistical effect, and large effects were found (direct effects). However, it may have influenced the size of the smaller indirect measurement effects. Thirdly, in four of the 14 classes, the random assignment on the second occasion did not prevent differences emerging between some of the subgroups on previous expectancy scores. When these classes were eliminated, both direct and indirect measurement effects remained essentially the same. In addition, multi-level analyses confirmed the presence of an indirect measurement effect. Fourthly, in the present study no comparison was made with children's expectancies concerning a non-alcoholic drink, which has recently been added to the measurement procedure by Query *et al.* (1998). Although this comparison could be an improvement, the results reported here are likely to be specific to alcohol, given mastery of the alcohol concept at an earlier age (e.g. Fossey, 1994), the high percentage of correct responses on the alcohol-identification questions here and the specificity to alcohol of expectancies reported by Query *et al.* (1998).

In conclusion, rather than recommending the future use of the questionnaire used here we recommend the careful development of similar expectancy instruments for use with children. Future instruments should contain an equal amount of positive and negative items, better cover the domain of children's expectancies and not use a puppet-reference. Recent promising examples may be found in Dunn & Goldman (1996) and in Query *et al.* (1998). However, given the results here, we would recommend to test these methods carefully as well as future instruments of children's expectancies for not producing unintended modifications of positive expectancies due their measurement.

Acknowledgements

This study was supported by a grant from SGO (Stimulation Program of Health and Health Care Research). We wish to thank Joyce van Reijen and Tischa Neve for data collection, Peter Molenaar, Don Mellenbergh, Harry Vorst, Fren Smulders, Robert West and Gerard van Breukelen for statistical advice and Ken Sher and Conor Dolan and the reviewers for helpful comments on earlier drafts of the manuscript. We would like to thank Mark Goldman for sending us the requested materials.

Notes

[1] In fact, the original CARE was derived from the AEQ-A (the adolescent form of the AEQ, Christiansen *et al.*, 1982) that contains one negative scale: cognitive and motor impairment. After a first administration of the CARE two scales were removed for reasons of poor internal consistency, including the only negative scale. The resulting 39-item CARE instrument only measures positive expectancies. This version was used by Kraus *et al.* (1994).

[2] To assess the model-fit, several measures can be used. Various indices of goodness of fit are reported, including the χ^2, in relation to the degrees of freedom, and its *p*-value (for an overview see Jöreskog, 1993). However, "statistical goodness-of-fit tests are often more a reflection on the size of the sample than on the adequacy of the model" (Browne & Cudeck, 1993). Therefore, relative fit-

measures are recommended such as the Root Mean Square Error of Approximation (RMSEA), which should be below 0.05 in case of good fit (Browne & Cudeck, 1993). In multi-group analyses its value should be multiplied with the square root of the number of groups ("cRMSEA"). Another popular index is the Comparative Fit Index (CFI, Bentler, 1990), which should approach 1 in case of good fit.

[3] For a comparison with the CARE, the number of items of the scale should be accounted for with the Spearman–Brown formula (e.g. Cronbach, 1990). For the comparison that is closest to the CARE (positive expectancies in the PP condition), this adjustment results in an estimated reliability of 0.77 for the positive scale in the 23-item model, and of 0.86 for the positive scale in the 14-item reduced model (test–retest 39-item CARE was 0.87; Kraus *et al.*, 1994).

References

BENTLER, P. M. (1990) Comparative fit indexes in structural models, *Psychological Bulletin*, 107, 238–246.

BROWN, S. A., GOLDMAN, M. S., INN, A. & ANDERSON, L. (1980) Expectations of reinforcement from alcohol: their domain and relation to drinking patterns, *Journal of Consulting and Clinical Psychology*, 48, 419–426.

BROWNE, M. W. & CUDECK R. (1993) Alternative ways of assessing model fit, in: BOLLEN, K. A. & SCOTT LONG, J. (Eds) *Testing Structural Equation Models*, pp. 136–162 (Newbury Park, Sage).

CASSWELL, S., BRASCH, P., GILMORE, L. & SILVA, P. (1985) Children's attitudes to alcohol and awareness of alcohol-related problems, *British Journal of Addiction*, 80, 191–194.

CECI, S. J. & BRUCK, M. (1993) Suggestibility of the child witness: a historical review and synthesis, *Psychological Bulletin*, 113, 403–439.

CHRISTIANSEN, B. A., GOLDMAN, M. S. & INN, A. (1982) Development of alcohol-related expectancies in adolescents: separating pharmacological from social-learning influences, *Journal of Consulting and Clinical Psychology*, 50, 336–344.

CHRISTIANSEN, B. A., SMITH, G. T., ROEHLING, P. V. & GOLDMAN, M. S. (1989) Using alcohol expectancies to predict adolescent drinking behaviour after one year, *Journal of Consulting and Clinical Psychology*, 57, 93–99.

CRONBACH, L. J. (1990) *Essentials of Psychological Testing*, 5th edn (New York, Harper & Row).

DUNN, M. E. & GOLDMAN, M. S. (1996) Empirical modelling of an alcohol expectancy memory network in elementary school children as a function of grade, *Experimental and Clinical Psychopharmacology*, 4, 209–217.

FOSSEY, E. (1994) *Growing Up With Alcohol* (London, Routledge).

FROMME, K., STROOT, E. & KAPLAN, D. (1993) Comprehensive effects of alcohol: development and psychometric assessment of a new expectancy questionnaire, *Psychological Assessment*, 5, 19–26.

GOLDMAN, M. S., BROWN, S. A. & CHRISTIANSEN, B.A. (1987) Expectancy theory: thinking about drinking, in: BLANE, H. T. & LEONARD, K. E. (Eds) *Psychological Theories of Drinking and Alcoholism*, pp. 181–226 (New York, Guilford Press).

HEDEKER, D. & GIBBONS, R. D. (1996) MIXREG: a computer progam for mixed-effects regression analysis with autocorrelated errors, *Computer Methods and Programs in Biomedicine*, 49, 229–252.

JOHNSON, E. O., ARRIA, A. M., BORGES, G., IALONGO, N. & ANTONY, J. C. (1995) The growth of conduct problem behaviours from middle childhood to early adolescence: sex differences and the suspected influence of early alcohol use, *Journal of Studies on Alcohol*, 56, 661–671.

JONES, B. T. & MCMAHON, J. (1994) Negative and positive alcohol expectancies as predictors of abstinence after discharge from a residential treatment programme: a one- and three-month follow-up study in males, *Journal of Studies on Alcohol*, 55, 543–548.

JÖRESKOG, K. (1993) Testing structural equation models, in: BOLLEN, K. A. & SCOTT LONG, J. (Eds) *Testing Structural Equation Models*, pp. 294–316 (Newbury Park, Sage).

JÖRESKOG, K. G. & SÖRBOM, D. (1993) *New Features in LISREL 8* (Chicago, Scientific Software International Inc.).

KRAUS, D., SMITH, G. T. & RATNER, H. H. (1990) The development of alcohol expectancies in children. Paper presented at the 1990 Annual Meeting of the Research Society on Alcoholism, Toronto, Ontario.

KRAUS, D., SMITH, G. T & RATNER, H. H. (1994) Modifying alcohol-related expectancies in grade-school children, *Journal of Studies on Alcohol*, 55, 535–542.

LEIGH, B. C. & STACY, A. W. (1993) Alcohol outcome expectancies: scale construction and predictive utility in higher order confirmatory factor models, *Psychological Assessment*, 5, 216–229.

MILLER, P. M., SMITH, G. T. & GOLDMAN, M. S. (1990) Emergence of alcohol expectancies in childhood: a possible critical period, *Journal of Studies on Alcohol*, 51, 343–349.

NOLL, R. B., ZUCKER, R. A. & GREENBERG, G. S. (1990) Identification by smell among pre-schoolers: evidence for early socialization about drugs occurring in the home, *Child Development*, 61, 1520–1527.

PEDERSEN, W. & SKONDAL, A. (1998) Alcohol consumption debut: predictors and consequences, *Journal of Studies on Alcohol*, 59, 32–42.

QUERY, L. R., ROSENBERG, H. & TISAK, M. S. (1998) The assessment of young children's expectancies of alcohol versus a control substance, *Addiction*, 93, 1521–1529.

WIERS, R. W. (1998) Bad expectations? Cognitive and neuropsychological indicators of enhanced risk for alcoholism, Doctoral Dissertation. Delft, The Netherlands, Eburon.

WIERS, R. W., GUNNING W. B. & SERGEANT, J. A. (1998) Do young children of alcoholics have more positive or more negative alcohol-related expectancies than controls? *Alcoholism: Clinical and Experimental Research*, 22, 1855–1863.

WIERS, R.W., HOOGEVEEN, K.J., SERGEANT, J.A. & GUNNING W.B. (1997) High and low dose expectancies and the differential associations with drinking in male and female adolescents and young adults, *Addiction*, 92, 871–888.